

# A formula for multiple classifiers in data mining based on Brandt semigroups

A.V. KELAREV,<sup>1</sup> J.L. YEARWOOD,<sup>2</sup> M.A. MAMMADOV<sup>3</sup>

School of Information Technology  
and Mathematical Sciences  
University of Ballarat  
P.O. Box 663, Ballarat  
Victoria 3353, Australia  
{a.kelarev,j.yearwood,m.mammadov}@ballarat.edu.au

## Abstract

A general approach to designing multiple classifiers represents them as a combination of several binary classifiers in order to enable correction of classification errors and increase reliability. This method is explained, for example, in Witten and Frank [44], Section 7.5. The aim of this paper is to investigate representations of this sort based on Brandt semigroups. We give a formula for the maximum number of errors of binary classifiers, which can be corrected by a multiple classifier of this type. Examples show that our formula does not carry over to larger classes of semigroups.

## 1 Introduction

Classification of data plays one of the central roles in data mining and in practical applications of artificial intelligence methods in general, see for example, [42] and [7, 8, 9, 16, 34, 43]. A well known method of designing efficient multiple classifiers consists in representing them as several binary classifiers combined in one scheme. This method is very effective, and it is often advisable to apply it even in situations where it is possible to build multiple classifiers analysing the data directly, see Witten and Frank [44], Section 7.5.

The main advantage of using combined multiple classifiers is that they can correct errors of individual binary classifiers and produce correct classifications despite individual classification errors. It is usually desirable to choose convenient representation for the class set of the multiple classifier and to ensure that it has a small set of generators.

The problem of finding the number of errors of individual binary classifiers that a multiple classifier can correct in general is rather complicated. It is well known that in full generality this problem is related to several other very difficult algorithmic problems, see [11] and [42].

This note uses semigroup rings to introduce additional structure on the class sets of multiple classifiers, which makes it possible to generate these sets with a small number of generators. Investigating properties

---

<sup>1</sup>The first author was supported by Australian Research Council, Discovery grant DP0449469.

<sup>2</sup>The second author was supported by Queen Elizabeth II Fellowship and ARC Discovery grant DP0211866.

<sup>3</sup>The third author has been supported by Discovery grant DP0556685 and Linkage grant LP0669752 from Australian Research Council.

of these multiple classifiers, first of all we are interested in inverse semigroups, which have been actively investigated by many authors (see, for example, [4, 5, 12, 13, 14, 15, 20, 21, 22, 23, 24, 25, 28, 36]).

An important subclass of inverse semigroups is that of Brandt semigroups. They have been explored, for example, in [17, 18, 33, 37]. Recall that a Brandt semigroup is an inverse completely 0-simple semigroup. The role of Brandt semigroups is illustrated by the well known fact that the principal factors of all inverse semigroups are Brandt semigroups. It is well known that every finite inverse semigroup can be ‘built’ from Brandt semigroups by means of ideal extensions. This is analogous to the method used for reduction of the the radical of an arbitrary semigroup algebra of a finite semigroup  $S$  to the radicals of group algebras of the subgroups of  $S$ , see [19].

It is remarkable that in the special case of Brandt semigroups and their subsemigroups we have managed to obtain a fairly concise and quite sophisticated formula for the number of errors of binary classifiers, which can be corrected by the corresponding multiple classifiers. This formula is the main result of our paper. It is presented in Theorem 2.

The proof of correctness of our formula is rather nontrivial. Examples are given to show that the formula does not directly generalize to all inverse semigroups and other classes of semigroups.

## 2 Motivation

For the standard notions of semigroup theory and classification methods used in data mining the reader is referred to [27, 30, 31, 34, 42, 44].

We are going to consider the problem of combining several binary classifiers into a multiple classifier. Suppose that there are binary classifiers  $b_1, \dots, b_m$ . Each of these classifiers divides their input data into two classes by producing an output 0 or 1 for each input element. If  $o_1, \dots, o_m$  are the outputs of the binary classifiers, then the sequence  $(o_1, \dots, o_m)$  is called a *class vector* of the combined multiple classifier, and the set of all class vectors is called the *class set*. Each class vector represents one class in the classification produced by the multiple classifier, see [44], Table 7.1.

First, let us briefly review basic essential properties required of the class sets. Let  $\mathbb{F} = GF(2)$  be the finite field of order two, i.e., the set  $\{0, 1\}$  with the standard addition and multiplication. Denote by  $\mathbb{F}^m$  the set of all sequences of all possible of outcomes of the binary classifiers which can occur in general. This means that

$$\mathbb{F}^m = \{(r_1, \dots, r_m) : r_1, \dots, r_m \in \mathbb{F}\}. \quad (1)$$

For every element  $r \in \mathbb{F}^m$ , denote by  $r_i$  the  $i$ -th component of the sequence  $r$ , so that

$$r = (r_1, \dots, r_m) \in \mathbb{F}^m. \quad (2)$$

The *weight*, or *Hamming weight*, of a sequence  $c$  in  $\mathbb{F}^m$  is denoted by  $\text{wt}(c)$ . It is defined as the number of nonzero coordinates in the vector  $c$ . The *weight* of a class set  $C$  is the minimum weight of a nonzero element in  $C$ . The *minimum distance* of a class set  $C$  is the minimum weight among all weights of nonzero differences between pairs of elements in  $C$ . If the class set  $C$  forms a linear subspace of  $\mathbb{F}^m$ , then its minimum distance always coincides with its weight.

For any real number  $x$ , denote by  $[x]$  the *integral part* of  $x$ , or the *floor* of  $x$ , that is the largest integer which does not exceed  $x$ . It is well-known and easy to verify that the number of errors of binary

classifiers, which the multiple classifier can correct, is equal to  $\lfloor (d-1)/2 \rfloor$ , where  $d$  is the minimum distance of the class set of the classifier.

The *information rate* of a class set  $C$  in  $\mathbb{F}^m$  can be defined as  $\log_2(|C|)/m$ . It reflects the proportion of output of the binary classifiers used to produce the outcomes of the multiple classification, as opposed to additional efforts spent on increasing reliability and correcting classification errors.

All sequences of the class set  $C$  can be written down in a matrix  $M$  to discuss their properties. If  $M$  has two identical columns, this means that two binary classifiers produce identical outputs. This duplication is very inefficient, even though it could help to correct classification errors. Therefore, in a situation like this, one of these classifiers can be removed and a better scheme can be devised. Likewise, it is undesirable to have strong correlation or functional dependencies between very small sets of columns in  $M$  or between binary classifiers.

According to [44], Section 7.5, for a classifier with a class set  $C$  to be efficient, the class  $C$  must satisfy the following most essential basic properties:

- (1) The minimum distance of  $C$  must be large.
- (2) The information rate of  $C$  must be large.
- (3) A convenient method of generating the set  $C$  is essential.
- (4) If all vectors of  $C$  are recorded in a matrix  $M$ , then there should not be strong correlation or functional dependencies between small sets of columns of  $M$ . In particular, the matrix  $M$  should not have duplicate columns.

As usual the standard addition is defined on  $\mathbb{F}^m$  componentwise, i.e., the sum of two arbitrary sequences  $(r_1, \dots, r_m)$  and  $(s_1, \dots, s_m)$  in  $\mathbb{F}^m$  is defined as

$$(r_1, \dots, r_m) + (s_1, \dots, s_m) = (r_1 + s_1, \dots, r_m + s_m). \quad (3)$$

### 3 Main definition

Instead of storing the whole large class set  $C$  in computer memory, it is convenient to be able to generate  $C$  with one or more generators. To this end we are going to introduce a multiplication on the set  $\mathbb{F}^m$ . It will allow us to multiply the generators with arbitrary elements of  $\mathbb{F}^m$  and to take their sums.

In order to generate classifiers with known properties and find optimal multiple classification schemes, we are going to take a finite semigroup  $S$  and use its algebra to introduce additional structure on the class set of a multiple classifier. The structure will enable us to find small generating sets for the classifier. The number of elements in  $S$  will have to be equal to the number of binary classifiers being combined. In other words, we assume that

$$S = \{s_1, \dots, s_m\}.$$

The *semigroup algebra*  $\mathbb{F}[S]$  is the set

$$\mathbb{F}[S] = \left\{ \sum_{s \in S} f_s s : f_s \in \mathbb{F} \right\}$$

with addition and multiplication defined by the associative and distributive laws and the rules

$$\sum_{s \in S} f_s s + \sum_{s \in S} f'_s s = \sum_{s \in S} (f_s + f'_s) s,$$

$$\left( \sum_{s \in S} r_s s \right) \left( \sum_{t \in S} r'_t t \right) = \sum_{s, t \in S} (r_s r'_t) st.$$

It is natural to regard  $S$  as being embedded in  $\mathbb{F}[S]$  by identifying each element  $s$  of  $S$  with  $1s$  in  $\mathbb{F}[S]$ . Every element of the form  $fs$ , where  $f \in \mathbb{F}$  and  $s \in S$ , is called a *homogeneous element* of  $\mathbb{F}[S]$ . Each term  $r_s s$  is called a *homogeneous component*, or the *s-component*, of the element  $r = \sum_{s \in S} r_s s$ . The *support* of an element  $r \in \mathbb{F}[S]$  is defined as the set

$$\text{supp}(r) = \{s \in S : r_s \neq 0\}.$$

The weight of  $r$  is equal to the cardinality of  $\text{supp}(r)$ . Let us refer to [30] and [38] for preliminaries on semigroup rings and other related constructions (see also [1, 2, 3, 6, 10, 15, 30, 31, 32]).

Regarded as an abelian group, the set  $\mathbb{F}^m$  is isomorphic to the additive group of the semigroup algebra  $\mathbb{F}[S]$ . In order to introduce an additional operation on the class set  $\mathbb{F}^m$ , let us identify the set  $\mathbb{F}^m$  with the semigroup algebra  $\mathbb{F}[S]$  by identifying each sequence  $r \in \mathbb{F}^m$  with the element  $\sum_{i=1}^m r_i s_i$  of the semigroup algebra  $\mathbb{F}[S]$ . This means that

$$(r_1, \dots, r_m) = r_1 s_1 + \dots + r_m s_m \in \mathbb{F}[S] = \mathbb{F}^m. \quad (4)$$

Given an arbitrary element  $r = r_1 s_1 + \dots + r_m s_m \in \mathbb{F}[S]$  and any  $1 \leq i \leq m$ , we introduce the notation  $r_{s_i} = r_i$ . It allows us to rewrite any element  $r \in \mathbb{F}^m = \mathbb{F}[S]$  as

$$r = \sum_{s \in S} r_s s. \quad (5)$$

Thus, the set  $\mathbb{F}^m = \mathbb{F}[S]$  has been endowed with a product defined, for  $x, y \in \mathbb{F}[S]$ , by the rule

$$\sum_{s \in S} x_s s \cdot \sum_{t \in S} y_t t = \sum_{s, t \in S} x_s y_t \cdot st. \quad (6)$$

Clearly, the *weight*  $\text{wt}(r)$  of an element

$$r = \sum_{s \in S} r_s s \in \mathbb{F}[S] = \mathbb{F}^m \quad (7)$$

coincides with the number of nonzero coefficients  $r_s$  in (7).

Now we can use two operations to generate classifiers. An element  $r \in \mathbb{F}^m$  is said to be *generated* by the elements  $g_1, \dots, g_k \in \mathbb{F}^m$  if it is the sum of multiples of these generators. Every generator is allowed to contribute several summands to this sum, and one has to write their multipliers on both sides, since the multiplication does not commute. This means that  $r$  is generated by the elements  $g_1, \dots, g_k$  if and only if there exist  $\ell_{j,i}, r_{j,i} \in \mathbb{F}^m \cup \{1\}$  such that

$$r = \sum_{j=1}^{m_1} \ell_{j,1} g_1 r_{j,1} + \dots + \sum_{j=1}^{m_k} \ell_{j,k} g_k r_{j,k}. \quad (8)$$

Next we discuss how one can simplify the expression (8). This is a standard step in generating ideals in semigroup rings. We include a concise explanation for convenience. As usual, by  $S^1 = S \cup \{1\}$  we denote the semigroup  $S$  with identity adjoined in the standard way.

By the definition of a semigroup algebra, the multipliers occurring in these sums can be rewritten, for  $i = 1, \dots, k$ , as follows:

$$\ell_{j,i} = \sum_{s \in S} \ell_s^{(j,i)} s, \quad r_{(j,i)} = \sum_{s \in S} r_s^{(j,i)} s, \quad (9)$$

where  $\ell_s^{(j,i)}, r_s^{(j,i)} \in \mathbb{F}$ . If we substitute (9) in (8) and use the associative and distributive laws, then we see that, in order to generate new elements, it is enough to multiply the generators by homogenous elements on the right and left and take sums of these products. Therefore to simplify the notation, we may assume that all the multipliers in (8) have been chosen as homogeneous elements from the very beginning, so that

$$\ell_{j,i} = \ell^{(j,i)} s^{j,i}, \quad r_{j,i} = r^{(j,i)} t^{j,i}, \quad (10)$$

for some  $\ell^{(j,i)}, r^{(j,i)} \in \mathbb{F}$  and  $s^{(j,i)}, t^{(j,i)} \in S$ . Since  $\mathbb{F} = GF(2) = \{0, 1\}$ , we see that all nonzero coefficients in the expression for the class set are equal to 1. Thus we can write

$$r = \sum_{j=1}^{m_1} s^{(j,1)} g_1 t^{(j,1)} + \dots + \sum_{j=1}^{m_k} s^{(j,k)} g_k t^{(j,k)}. \quad (11)$$

Accordingly, the whole class set  $C$  of a multiple classifier is said to be *generated* by the elements  $g_1, \dots, g_k$  in  $\mathbb{F}^m$  if  $C$  coincides with the set of all sums of multiples of these generators, i.e.,

$$\begin{aligned} C &= C(g_1, \dots, g_k) \\ &= \left\{ \sum_{j=1}^{m_1} s^{(j,1)} g_1 t^{(j,1)} + \dots + \sum_{j=1}^{m_k} s^{(j,k)} g_k t^{(j,k)} : \text{where } s^{(j,i)}, t^{(j,i)} \in S^1 \right\}. \end{aligned} \quad (12)$$

In this case the notation  $C = C(g_1, \dots, g_k)$  is used when it is necessary to indicate the generators explicitly.

It is easy to show that all maximum weights of class sets can be always achieved by generating the sets with single generators. On the other hand, considering several generators makes it possible to generate class sets with larger information rates.

## 4 Preliminaries on Brandt semigroups

For convenience of the readers, this section includes concise preliminaries on Brandt semigroups, following [27]. As usual, to avoid possible ambiguities that may occur in considering semigroup rings, zeros of semigroups will be denoted by  $\theta$ , and the symbol 0 will denote the zero of a ring. As customary, the sets  $\{0\}$  and  $\{\theta\}$  will be also denoted by 0 and  $\theta$ , respectively.

Suppose that  $G$  is a group,  $I$  and  $\Lambda$  are nonempty sets,  $e$  the identity of  $G$ , and  $P = [p_{\lambda i}]$  is a  $(\Lambda \times I)$ -matrix with entries  $p_{\lambda i} \in G$ , for all  $\lambda \in \Lambda$ ,  $i \in I$ . As usual, we denote by  $G^1 = G \cup \{1\}$  and  $G^0 = G \cup \{0\}$  the group  $G$  with identity 1 and, respectively, zero  $\theta$  adjoined in a standard fashion.

The *Rees matrix semigroup*  $M^0(G; I, \Lambda; P)$  over  $G$  with *sandwich-matrix*  $P$  consists of all triples  $(g; i, \lambda)$ , where  $i \in I$ ,  $\lambda \in \Lambda$ , and  $g \in G^0$ , where all triples  $(\theta; i, \lambda)$  are identified with  $\theta$ , and where multiplication is defined by the rule

$$(g_1; i_1, \lambda_1)(g_2; i_2, \lambda_2) = (g_1 p_{\lambda_1 i_2} g_2; i_1, \lambda_2).$$

A *Brandt semigroup* is a completely 0-simple inverse semigroup. It is well known that a semigroup  $S$  is a Brandt semigroup if and only if it can be represented as a Rees matrix semigroup  $M = M^0(G; I, \Lambda; P)$  over a group  $G$  with zero such that  $|I| = |\Lambda|$  and  $P$  is the identity matrix, see [27]. We will always use this representation for the Brandt semigroups, and identify the elements of the sets  $I$  and  $\Lambda$  so that  $I = \Lambda$  and

$$p_{i\lambda} = \begin{cases} e & \text{if } i = \lambda, \\ \theta & \text{otherwise,} \end{cases} \quad (13)$$

where  $e$  is the identity of the group  $G$ . Following [17], [18], [33], and [37], we also denote the Brandt semigroup  $M(G; I, \Lambda; P)$  by  $B(G, |I|)$  or  $B(G, I)$ .

Let  $G$  be a group,  $M = M^0(G; I, \Lambda; P)$ , and let  $i \in I$ ,  $\lambda \in \Lambda$ . Put

$$\begin{aligned} G_{*\lambda} &= \{(g; i, \lambda) : g \in G, i \in I\} \cup \{\theta\}, \\ G_{i*} &= \{(g; i, \lambda) : g \in G, \lambda \in \Lambda\} \cup \{\theta\}, \\ G_{i\lambda} &= \{(g; i, \lambda) : g \in G\} \cup \{\theta\}. \end{aligned}$$

The following properties are very well known (see, for example, [27]).

**Lemma 1** *Let  $G$  be a group, and let  $M = M^0(G; I, \Lambda; P)$  be a completely 0-simple semigroup. Then, for all  $i, j \in I$ ,  $\lambda, \mu \in \Lambda$ ,*

- (i)  $G_{*\lambda}$  is an  $\mathcal{L}$ -class of  $M$  and a minimal nonzero left ideal of  $M$ ;
- (ii)  $G_{i*}$  is an  $\mathcal{R}$ -class of  $M$  and a minimal nonzero right ideal of  $M$ ;
- (iii)  $G_{i\lambda}$  is an  $\mathcal{H}$ -class of  $M$ , a left ideal of  $G_{i*}$  and a right ideal of  $G_{*\lambda}$ ;
- (iv)  $|G_{i\lambda}| = |G_{j\mu}|$ ;
- (v) each maximal subgroup of  $M$  coincides with  $G_{j\mu}$ , for some  $j \in I$ ,  $\mu \in \Lambda$ ;
- (vi) if  $p_{\lambda i} \neq \theta$ , then  $G_{i\lambda}$  is a maximal subgroup of  $M$  isomorphic to  $G$ ;
- (vii) if  $p_{\lambda i} = \theta$ , then  $G_{i\lambda}^2 = \theta$ ;
- (viii) every  $\mathcal{L}$ -class of  $M$  contains at least one maximal subgroup,  $G_{j\mu}$ ;
- (ix) every  $\mathcal{R}$ -class of  $M$  contains at least one maximal subgroup,  $G_{j\mu}$ .

## 5 Main results

Let  $S$  be a subsemigroup of the Brandt semigroup  $B = B(G; I)$ . For any  $i, \lambda \in I$ , put

$$S_{i\lambda} = S \cap G_{i\lambda}, \quad (14)$$

$$S_{*\lambda} = S \cap G_{*\lambda}, \quad (15)$$

$$S_{i*} = S \cap G_{i*}. \quad (16)$$

Let us introduce subsets

$$L = \{i \in I : S_{i*} = \theta\}, \quad (17)$$

$$R = \{\lambda \in \Lambda : S_{*\lambda} = \theta\}. \quad (18)$$

This means that  $L$  is the set of indices of empty or zero rows of  $S$ , and  $R$  is the set of indices of empty or zero columns. Putting

$$S_{*L} = \cup_{\lambda \in L} S_{*\lambda}, \quad (19)$$

$$S_{R*} = \cup_{i \in R} S_{i*}, \quad (20)$$

we see that  $S_{*L}$  is the set of columns of  $S$  corresponding to empty or zero rows. Similarly,  $S_{R*}$  is the set of rows corresponding to empty or zero columns of  $S$ . It follows from the definition of a Brandt semigroup, (13), (17), and (19) that  $S_{*L}$  coincides with the *left annihilator*  $\text{Ann}_\ell(S)$  of  $S$ , i.e.,

$$S_{*L} = \text{Ann}_\ell(S) = \{x \in S : xS = \theta\}. \quad (21)$$

Likewise,  $S_{R*}$  coincides with the *right annihilator*  $\text{Ann}_r(S)$  of  $S$ , i.e.,

$$S_{R*} = \text{Ann}_r(S) = \{x \in S : Sx = \theta\}. \quad (22)$$

The following theorem gives us a formula for the maximum error correcting capabilities of the classifiers  $C(g_1, \dots, g_k)$  in  $\mathbb{F}[S]$ .

**Theorem 2** *Let  $S$  be a subsemigroup of the Brandt semigroup  $B(G, I)$ , and let  $L$  and  $R$  be the sets introduced above. Denote by  $N$  the following maximum*

$$N = \max\{M_Z, M_L, M_R, M_G\}, \quad (23)$$

where

$$M_Z = |S_{R*} \cap S_{*L}|, \quad (24)$$

$$M_L = \max\{|S_{i*} \cap S_{*L}| : \text{for all } i \notin L\}, \quad (25)$$

$$M_R = \max\{|S_{*\lambda} \cap S_{R*}| : \text{for all } \lambda \notin R\}, \quad (26)$$

$$M_G = \max\{|S_{i\lambda}| : \text{for all } i \notin L, \lambda \notin R\}. \quad (27)$$

Then the maximum number  $E$  of errors of binary classifiers which can be corrected by a multiple classifier of the form  $C(g_1, \dots, g_k)$  in  $\mathbb{F}[S]$  is equal to

$$E = \left\lfloor \frac{N-1}{2} \right\rfloor. \quad (28)$$

This theorem does not generalise to arbitrary inverse semigroups, as the following example shows. Recall, that a *semilattice* is a commutative semigroup entirely consisting of idempotents.

**Example 3** Let  $k$  be a positive integer, and let  $Y = \{\theta, y_1, \dots, y_{2k}, \varepsilon\}$  be the semilattice with zero  $\theta$ , identity  $\varepsilon$  and all products  $y_i y_j$  defined by the rule

$$y_i y_j = \begin{cases} y_i & \text{if } i = j, \\ \theta & \text{otherwise,} \end{cases} \quad (29)$$

for all  $1 \leq i, j \leq 2k$ . Put

$$g = \theta + \sum_{i=1}^{2k} y_i + \varepsilon \in \mathbb{F}[Y].$$

Then it is routine to verify that the class set  $C(g)$  generated by  $g$  coincides with the set  $\mathbb{F}g$ . Therefore, the minimum distance of  $C(g)$  is equal to  $2k + 2$ , and so we see that it can correct  $k$  errors of binary classifiers. However,  $\text{Ann}_\ell(Y) = \text{Ann}_r(Y) = \theta$  and all subgroups of  $Y$  are singletons. Therefore  $M_Z = M_L = M_R = M_G = 1$ .

Our next example shows that the main theorem does not generalise to other classes of semigroups either. A *band* is a semigroup entirely consisting of idempotents. Bands regularly occur in solutions to various problems (see [26] for a recent example). A band  $B$  is said to be *rectangular* if it satisfies the identity  $xyx = x$ , for all  $x, y \in B$ .

**Example 4** Let  $k$  be a positive integer,  $I$  and  $\Lambda$  sets with  $|I| = |\Lambda| = 2k$ ,  $G = \{e\}$  a group of order one with identity  $e$ , and let  $P$  be the  $\Lambda \times I$ -sandwich matrix with  $p_{\lambda i} = e$  for all  $i \in I$ ,  $\lambda \in \Lambda$ . Then  $B = M^0(G; I, \Lambda; P)$  is a rectangular band with zero adjoined. Let

$$g = \sum_{i \in I} \sum_{\lambda \in \Lambda} (e; i, \lambda) \in \mathbb{F}[B].$$

Then it is routine to verify that the class set  $C(g)$  coincides with the set  $\mathbb{F}g$ . Therefore, the minimum distance of  $C(g)$  is equal to  $4k^2$ , and so it can correct  $2k^2 - 1$  errors of binary classifiers. However, there are no empty or zero rows in  $B$ , and we see that  $M_Z = M_L = M_R = M_G = 1$ .

## 6 Proofs

Our main theorem follows from a formula for the largest weights of class sets  $C(g_1, \dots, g_k)$  in  $\mathbb{F}[S]$ , which we record as an auxiliary proposition here. Notice that  $2 \left\lfloor \frac{x+1}{2} \right\rfloor$  is equal to the smallest even integer which is not less than  $x$ , and  $2 \left\lfloor \frac{x}{2} \right\rfloor$  is equal to the largest even integer which is less than  $x$ .

**Proposition 5** *In the conditions of Theorem 2 the largest weight  $W$  of the class set of a multiple classifier of the form  $C(g_1, \dots, g_k)$  in  $\mathbb{F}[S]$  is equal to*

$$W = 2 \left\lfloor \frac{N+1}{2} \right\rfloor. \quad (30)$$



*Proof.* Let  $C = C(g_1, \dots, g_k)$  be the class set which has the largest weight  $W$  among all class sets of this form in  $\mathbb{F}[S]$ . Evidently, all class sets  $C(g_1), \dots, C(g_k)$  are subsets of  $C = C(g_1, \dots, g_k)$ . The weight of a subset is never less than the weight of the superset. Therefore all the class sets  $C(g_1), \dots, C(g_k)$  have the same weight  $W$ . Hence it suffices to prove the proposition in the case where  $k = 1$ ,  $C$  is generated by one element  $g = g_1$ , so that  $C = C(g)$  and  $W = \text{wt}(C)$ .

Let us divide the proof of equality (30) into two parts.

**Part 1.** We are going to prove the inequality

$$\text{wt}(C(g)) \geq 2 \left\lfloor \frac{N+1}{2} \right\rfloor. \quad (31)$$

By the definition of  $N$ , it suffices to verify four inequalities:

$$\text{wt}(C(g)) \geq 2 \left\lfloor \frac{M_Z+1}{2} \right\rfloor, \quad (32)$$

$$\text{wt}(C(g)) \geq 2 \left\lfloor \frac{M_L+1}{2} \right\rfloor, \quad (33)$$

$$\text{wt}(C(g)) \geq 2 \left\lfloor \frac{M_R+1}{2} \right\rfloor, \quad (34)$$

$$\text{wt}(C(g)) \geq 2 \left\lfloor \frac{M_G+1}{2} \right\rfloor. \quad (35)$$

The maximality of  $W = \text{wt}(C(g))$  shows that these inequalities will follow if we demonstrate that  $\mathbb{F}[S]$  contains elements  $g_Z, g_L, g_R$ , and  $g_G$  which generate the class sets satisfying

$$\text{wt}(C(g_Z)) = 2 \left\lfloor \frac{M_Z+1}{2} \right\rfloor, \quad (36)$$

$$\text{wt}(C(g_L)) = 2 \left\lfloor \frac{M_L+1}{2} \right\rfloor, \quad (37)$$

$$\text{wt}(C(g_R)) = 2 \left\lfloor \frac{M_R+1}{2} \right\rfloor, \quad (38)$$

$$\text{wt}(C(g_G)) = 2 \left\lfloor \frac{M_G+1}{2} \right\rfloor. \quad (39)$$

Hence in Part 1 of our proof it remains to find elements  $g_Z, g_L, g_R$ , and  $g_G$  with required properties.

First, we are going to find an element  $g_Z \in \mathbb{F}[S]$  satisfying (36). Put  $Z = S_{R*} \cap S_{*L}$ . The case where  $Z = \emptyset$  is trivial, and so we assume that  $M_Z = |Z| > 0$ . Consider the element

$$g_Z = \left( \sum_{x \in Z} x \right) + M_Z \theta \in \mathbb{F}[S]. \quad (40)$$

If  $M_Z$  is odd, then  $M_Z \theta = \theta$  in  $\mathbb{F}[S]$ . In this case we see that  $\text{wt}(g_Z) = |Z| + 1 = M_Z + 1 = 2 \left\lfloor \frac{M_Z+1}{2} \right\rfloor$ . On the other hand, if  $M_Z$  is even, then  $M_Z \theta = 0$  in  $\mathbb{F}[S]$ . Hence  $\text{wt}(g_Z) = M_Z = 2 \left\lfloor \frac{M_Z+1}{2} \right\rfloor$  in this case. Thus, we see that the weight of  $g_Z$  always satisfies

$$\text{wt}(g_Z) = 2 \left\lfloor \frac{M_Z+1}{2} \right\rfloor. \quad (41)$$

It follows from (21), (22) and (40) that  $\theta g_Z = g_Z \theta = 2M_Z \theta = 0$  and  $SZ = ZS = \theta$ . Therefore it is straightforward to verify that  $\mathbb{F}[S]g_Z = g_Z \mathbb{F}[S] = 2M_Z \theta = 0$ . We see that the class set  $C(g_Z)$  coincides with the linear space  $\mathbb{F}g_Z$  spanned by  $g_Z$  in  $\mathbb{F}[S]$ . Hence the weight of  $C(g_Z)$  is equal to the weight  $\text{wt}(g_Z)$ . Equality (41) shows that  $g_Z \in \mathbb{F}[S]$  satisfies (36), as required.

Second, we are going to find an element  $g_L \in \mathbb{F}[S]$  satisfying (37). The case where  $S_{(I \setminus L)*} \cap S_{*L} = \emptyset$  is trivial, and so we assume that  $M_L > 0$ . Choose  $i \in I \setminus L$  such that  $|S_{i*} \cap S_{*L}| = M_L$ . Put  $Z_i = S_{i*} \cap S_{*L}$  so that  $|Z_i| = M_L$ . Consider the element

$$g_L = \left( \sum_{x \in Z_i} x \right) + M_L \theta. \quad (42)$$

Take any nonzero element  $x$  with minimal weight in  $C(g_L)$ . Since  $Z_i \subseteq \text{Ann}_\ell(S)$ , expression (12) for  $C(g_L)$  simplifies as follows

$$C(g_L) = \left\{ \sum_{j=1}^{m_1} s_j g_L : \text{for some } s_j \in S^1 \right\}. \quad (43)$$

Therefore  $x$  can be recorded as

$$x = \sum_{j=1}^{m_1} s_j g_L, \quad (44)$$

for some  $s_1, \dots, s_{m_1} \in S \cup \{1\}$ . We may assume that all summands in (44) are nonzero and that the likely terms have been combined.

If  $1 \neq s_j \in S$  for some  $1 \leq j \leq m_1$ , then it follows from the definition of a Brandt semigroup,  $Z_i \subseteq S_{i*}$  and  $s_j g_L \neq 0$ , that  $s_j \in S_{*i}$ . Therefore the homogeneous components of different summands  $s_j g_L$  and  $s_k g_L$  never cancel, for any  $j \neq k$ . This implies that  $\text{wt}(x) = \sum_{j=1}^{m_1} \text{wt}(s_j g_L)$ . By the minimality of  $\text{wt}(x)$ , we get  $m_1 = 1$  and  $x = s_1 g_L$  for some  $s_1 \in S \cup \{1\}$ . It is straightforward to verify that  $\text{wt}(s_1 g_L) = \text{wt}(g_L)$ .

If  $|Z_i| = M_L$  is even, then  $M_L \theta = 0$ , and so  $\text{supp}(g_L) = Z_i$ . Hence  $\text{wt}(g_L) = M_L = 2 \left\lfloor \frac{M_L+1}{2} \right\rfloor$ , as required in (37). On the other hand, if  $M_L$  is odd, then  $M_L \theta = \theta$  and  $2 \left\lfloor \frac{M_L+1}{2} \right\rfloor = M_L + 1$ . Then we get  $\text{wt}(g_L) = M_L + 1 = 2 \left\lfloor \frac{M_L+1}{2} \right\rfloor$ . Thus  $g_L$  always satisfies (37).

Third, we are going to find an element  $g_R \in \mathbb{F}[S]$  satisfying (38). The case where  $S_{*(\Lambda \setminus R)} \cap S_{R*} = \emptyset$  is trivial, and so we assume that  $M_R > 0$ . Let us take  $\lambda \in \Lambda \setminus R$  such that  $|S_{*\lambda} \cap S_{R*}| = M_R$ . Put  $Z_\lambda = S_{*\lambda} \cap S_{R*}$  so that  $|Z_\lambda| = M_R$ . Consider the element

$$g_R = \left( \sum_{x \in Z_\lambda} x \right) + M_R \theta. \quad (45)$$

We claim that  $C(g_R)$  satisfies (38). The proof of this fact is dual to the prove given above for  $g_L$ , and so we omit it.

Fourth, we are going to find an element  $g_G \in \mathbb{F}[S]$  satisfying (39). The case where  $S_{I \setminus L, \Lambda \setminus R} = \emptyset$  is trivial, and so we assume that  $M_G > 0$ . Choose  $i \in I \setminus L$  and  $\lambda \in \Lambda \setminus R$  such that  $|S_{i\lambda}| = M_G$ . Consider

the element

$$g_G = \sum_{x \in S_{i\lambda}} x. \quad (46)$$

Pick any nonzero element  $x$  of minimal weight in the class set  $C(g_G)$ . By (12),

$$x = \left( \sum_{j=1}^{m_1} s_j g_G t_j \right) + M_G \theta, \quad (47)$$

for some  $s_j, t_j \in S^1$ . We may assume that all summands  $s_j g_G t_j$  in (47) are nonzero, and all similar terms have been combined.

Lemma 1 shows us that, for each  $j = 1, \dots, m_1$ , there exist  $i_j \in I$  and  $\lambda_j \in \Lambda$  such that  $s_j S_{i\lambda} t_j \subseteq S_{i_j \lambda_j}$ . Since  $G$  is a group, it follows from the definition of a Brandt semigroup  $M^0(G; I, \Lambda; P)$  that  $|s_j S_{i\lambda} t_j| = |S_{i\lambda}|$ , for all  $i = 1, \dots, m_1$ . The maximality of  $|S_{i\lambda}|$  implies  $|s_j S_{i\lambda} t_j| = |S_{i_j \lambda_j}|$ . It follows that

$$\text{supp}(s_j g_G t_j) \cap \text{supp}(s_k g_G t_k) \subseteq \{\theta\}, \quad (48)$$

whenever  $1 \leq j \neq k \leq m_1$ . Therefore, for every  $1 \leq j \leq m_1$ , we get

$$\text{wt}(x) \geq \text{wt}(s_j g_G t_j). \quad (49)$$

It also follows from the definition of a Brandt semigroup that

$$\text{wt}(s_j g_G t_j) = \text{wt}(g_G). \quad (50)$$

The minimality of  $\text{wt}(x)$  implies that  $\text{wt}(C(g_G)) = \text{wt}(x) = \text{wt}(g_G)$ .

If  $M_G$  is odd, then  $M_G \theta = \theta$  and  $\text{wt}(g_G) = M_G + 1 = \left\lfloor \frac{M_G + 1}{2} \right\rfloor$ . If, however,  $M_G$  is even, then  $M_G \theta = 0$  and so  $\text{wt}(g_G) = M_G = \left\lfloor \frac{M_G + 1}{2} \right\rfloor$ . Therefore  $g_G$  satisfies (39), as required.

Thus we have found the desired elements  $g_Z, g_L, g_R$  and  $g_G$ . This establishes that (31) holds.

**Part 2.** Now we are going to prove the reversed inequality

$$\text{wt}(C(g)) \leq W, \quad (51)$$

where  $g$  has been chosen in  $\mathbb{F}[S]$  as above so that the weight  $\text{wt}(C(g))$  achieves the largest possible value.

Consider a nonzero element  $x$  of minimal weight in  $C(g)$ . Since  $x \in C(g)$ , we get  $C(x) \subseteq C(g)$ , whence  $\text{wt}(C(x)) \geq \text{wt}(C(g))$ . Therefore in Part 2 of the proof we can replace  $g$  by  $x$ . To simplify further notation we may assume that the element  $g$  has been chosen from the very beginning so that

$$\text{wt}(C(g)) = \text{wt}(g). \quad (52)$$

The element  $g$  has a unique representation of the form  $g = \sum_{s \in S} g_s s$ . If  $\sum_{s \in S} g_s \neq 0$ , then  $\theta = (\sum_{s \in S} g_s) = g\theta \in C(g)$ , and so  $\text{wt}(C(g)) = 1 \leq W$ , as required. Further, in Part 2 we assume that

$$\sum_{s \in S} g_s = 0. \quad (53)$$

Clearly, this is equivalent to assuming that  $|\text{supp}(g)|$  is even.

The rest of Part 2 will be divided into several possible cases. It is straightforward that one of these cases always occurs.

**Case 1:**  $\text{supp}(g) \subseteq (S_{*L} \cap S_{R*}) \cup \{\theta\}$ . Put  $Z = S_{*L} \cap S_{R*}$  so that  $|Z| = M_Z$ .

In this case conditions (21) and (22) imply that  $S \text{supp}(g) \cup \text{supp}(g)S \subseteq \{\theta\}$ . This and (12) yield that the class set  $C(g)$  is contained in the linear space  $\mathbb{F}g + \mathbb{F}\theta$  spanned by  $g$  and  $\theta$  in  $\mathbb{F}[S]$ . It follows from this inclusion, as well as from (52), that  $\text{wt}(C(g)) \leq |Z| + 1$ .

If  $\text{supp}(g) \not\supseteq Z$ , then we get  $\text{wt}(C(g)) \leq |Z| \leq 2 \left\lfloor \frac{M_Z+1}{2} \right\rfloor \leq W$ . Hence (51) is satisfied.

Now suppose that  $\text{supp}(g) \supseteq Z$ . Then (53) implies that  $g = g_Z$ . It has been verified in Part 1 of our proof, that equality (41) holds for  $\text{wt}(g_Z)$ . This implies that (51) is satisfied again. Thus we see that in Case 1 the equality (51) always holds.

**Case 2:**  $\text{supp}(g) \subseteq S_{*L} \cup \{\theta\}$ , but  $\text{supp}(g) \not\subseteq S_{R*} \cup \{\theta\}$ . Then there exists  $s \in S_{*L} \setminus S_{R*}$  such that  $s \in \text{supp}(g)$ , and so  $g_s \neq 0$ . Clearly,  $s \in S_{j'\lambda}$  for some  $j' \in I \setminus R$  and  $\lambda \in L$ .

Since  $j' \notin R$  and  $S_{R*} = \text{Ann}_r(S)$  by (22), we see that there exists  $u \in S$  such that  $us \neq \theta$ . It follows that  $ug \neq 0$ , because the sandwich matrix of the Brandt semigroup is an identity matrix. Clearly,  $u \in S_{i\mu}$  for some  $i \in I$ ,  $\mu \in \Lambda$ . Put  $Z_i = S_{i*} \cap S_{*L}$ . Lemma 1 shows that  $\text{supp}(ug) \subseteq Z_i \cup \{\theta\}$ .

Obviously,  $ug \in C(g)$  implies  $C(ug) \subseteq C(g)$ ; whence  $\text{wt}(C(ug)) \geq \text{wt}(C(g))$ . By the maximality of  $\text{wt}(C(g))$ , we get  $\text{wt}(C(ug)) = \text{wt}(C(g))$ . Hence we could have chosen  $ug$  instead of  $g$  in the beginning of this part of the proof. To simplify the notation we will assume that  $g$  satisfies an additional property which  $ug$  possesses. Namely, since  $\text{supp}(ug) \subseteq Z_i$ , we may assume that  $\text{supp}(g) \subseteq Z_i \cup \{\theta\}$  from the very beginning.

If  $\text{supp}(g) \not\supseteq Z_i$ , then it is clear that  $\text{wt}(g) \leq |Z_i| \leq 2 \left\lfloor \frac{M_L+1}{2} \right\rfloor \leq W$ . Therefore (52) implies  $\text{wt}(C(g)) \leq W$ . Hence (51) is satisfied in this situation.

Now consider the subcase where  $\text{supp}(g) \supseteq Z_i$ . Then (53) implies that  $g = g_L$ . We have already shown in Part 1, that  $\text{wt}(C(g_L)) = \text{wt}(g_L) = M_L$ . Therefore (52) implies that (51) holds again. Thus in Case 2 the inequality (51) is satisfied too.

**Case 3:**  $\text{supp}(g) \subseteq S_{R*}$  but  $\text{supp}(g) \not\subseteq S_{*L}$ . This case is dual to Case 2, and so a dual argument shows that (51) holds true.

**Case 4:**  $\text{supp}(g) \not\subseteq S_{R*}$  and  $\text{supp}(g) \not\subseteq S_{*L}$ . Then there exists  $s \in \text{supp}(g) \cap (S \setminus S_{R*}) \cap (S \setminus S_{*L})$ . Clearly,  $s \in S_{j'\lambda'}$  for some  $j' \in I \setminus R$  and  $\lambda' \in \Lambda \setminus L$ .

Notice that  $j' \notin R$ ,  $\lambda' \notin L$  and  $S_{R*} = \text{Ann}_r(S)$ ,  $S_{*L} = \text{Ann}_\ell(S)$  by (22) and (21), respectively. Therefore it follows from the definition of a Brandt semigroup that there exist  $u, v \in S$  such that  $usv \neq \theta$ . Hence  $ugv \neq 0$ . We can find  $i, j \in I$  and  $\lambda, \mu \in \Lambda$  such that  $u \in S_{i\mu}$  and  $v \in S_{j\lambda}$ . Then Lemma 1 establishes that  $\text{supp}(ugv) \subseteq S_{i\lambda}$ .

As before,  $ugv \in C(g)$  implies  $C(ugv) \subseteq C(g)$  and  $\text{wt}(C(ugv)) \geq \text{wt}(C(g))$ . By the choice of  $g$ , we get  $\text{wt}(C(ugv)) = \text{wt}(C(g))$ . Therefore to simplify the notation we may assume that  $ugv$  has been chosen as  $g$  from the very beginning. This means that further we can use the inclusion  $\text{supp}(g) \subseteq S_{i\lambda} \cup \{\theta\}$ .

If  $\text{supp}(g) \not\supseteq S_{i\lambda}$ , then it follows that  $\text{wt}(g) \leq |S_{i\lambda}| \leq 2 \left\lfloor \frac{M_G+1}{2} \right\rfloor \leq W$ . By (52), we get  $\text{wt}(C(g)) \leq W$ . Therefore (51) holds true. Hence it remains to consider the subcase, where  $\text{supp}(g) \supseteq S_{i\lambda}$ . Then (53) implies that  $g = g_G$ . It has already been established in Part 1, that (39) holds for  $\text{wt}(g_G)$ . By (52), we get (51) again. Thus in Case 4 the inequality (51) always holds. This completes the proof of the proposition.  $\square$

*Proof of Theorem 2.* Proposition 5 shows that  $W$  satisfies (30). We know that every class set with minimum distance  $W$  can correct  $\left\lfloor \frac{W-1}{2} \right\rfloor$  errors of binary classifiers. If  $N$  is even, then  $W = N$ , and the number  $E$  of errors corrected by  $C$  satisfies (28), as required in our theorem. On the other hand, if  $N$  is odd, then  $W = N + 1$  and the number of errors corrected by  $C$  is given by  $E = \left\lfloor \frac{W-1}{2} \right\rfloor = \left\lfloor \frac{N}{2} \right\rfloor = \left\lfloor \frac{N-1}{2} \right\rfloor$ , as required by (28) again. This completes the proof of our theorem.  $\square$

## 7 Example of an Application

Using multiple classifiers to achieve higher accuracy in prediction is quite common in the literature. Following [42], in this section we consider some examples from data classification where multiple classifiers are used or can be used.

**1.** Neural Networks are one of the most successful techniques used in data classification. For example, Feedforward Neural Networks (FNN) have been applied to classification problems from different areas, see [34].

A multilayer feedforward neural network can have any number of layers and any number of units (neurons) per layer. The first layer is called the input layer and the last layer is called the output layer. The middle layers are called hidden layers. Each neuron-to-neuron connection is modified by a weight or connection strength. In addition, each neuron has an extra input that is assumed to have a constant value of one, and the weight that modifies this extra input is called the bias. In FNN all information propagates along the connections in the direction of network inputs to network outputs, hence the term feedforward.

A network is trained with inputs as well as outputs from training data or sample data. Then the network learns the mapping from inputs to corresponding outputs. This is called supervised learning. In order for the network to learn the patterns of the data, a learning algorithm is needed. Backpropagation is the learning algorithm most commonly used for the feedforward neural networks.

Backpropagation is an algorithm based on least squares that modifies network weights to minimise the mean squared errors between the predicted and actual outputs of the network. Backpropagation is a supervised learning algorithm. Once the network is trained, the weights are optimized and these optimized weights can then be used to compute outputs for new inputs.

Backpropagation tries to minimise the sum of squared errors, by forcing the network weights to change in such a way that errors are minimised. Backpropagation training consists of three steps:

1. **Output Calculations:** present the given input vector to the network inputs and run the network: compute the activation functions sequentially forward from the first hidden layer to the output layer.

2. **Error Backpropagation:** compute the difference between the predicted output and the actual output. Propagate the error sequentially backward from the output layer to the input layer.
3. **Weight Modification:** for every connection, change the weight by modifying that connection in proportion to the error.

When these three steps have been performed for every example from the data, one epoch of training has occurred. Learning usually runs through thousands of epochs, either until a predetermined maximum number of epochs is reached, or the network output error falls below an acceptable threshold. Training (learning) of a network can be time consuming, depending on the network size, size of the training data set, number of epochs, and the desired network output errors.

Backpropagation may lead the weights in a neural network to values that correspond to a local minimum of the error function. This local minimum may be substantially different from the global minimum that corresponds to the best choice of weights. Therefore, the performance of a single neural network based on such a solution may not be sufficiently high. In order to overcome this drawback, it is common in the literature to train neural networks many times (usually hundreds of times) and to generate a set of neural networks. The prediction corresponding to the set of networks is defined as an average of the outputs. It turns out that the performance of such a set of neural networks (that is, multiple classifiers) is significantly higher than each of these single classifiers.

There are some studies in the literature that aim to find the global optimal solution of the error function and then to generate one single neural network. The performance of this network is expected to be higher than the networks based on solutions far away from the global solution. However, the calculations in [41] showed that, the set of neural networks (1000 trials) based on different local optimal solutions provided more accurate predictions compared a single network based on better (that is, global) solution.

The outputs of neural networks can be presented in the form of binary values. In this case, we can define the prediction of multiple neural networks on the basis of these binary values rather than average outputs. This may decrease the impact of very “bad” solutions that usually occur during the process of multiple training.

2. Many classification algorithms use some parameters whose values are hard to determine for a given dataset. Taking different values for these parameters and generating multiple classifiers might be more efficient in this case. The accuracy of prediction in this case might be much higher than if we apply one single classifier spending time for tuning the parameters in an “optimal” way.

For example, SVM, together with some parameters, uses different kernels. Instead of trying to choose the more appropriate kernel, we can use different kernels and generate different classifiers. We can expect that the combination of these classifiers will provide more accurate prediction.

3. BoosTexter is a well known algorithm used in multi-label classification problems, see [39]. The purpose of boosting is to find an accurate classification rule by combining many weak or base hypotheses, each of which may be only moderately accurate. Note that, these weak hypotheses are in fact simple binary classifiers. Therefore, we can say that BoosTexter is based on the idea of combining multiple single classifiers. This, in particular, shows how simple binary classifiers can be used to handle datasets with multi-classes or multi-label classes.

4. The algorithm developed in [35] applies three different distance measures that leads to three

different versions of this algorithm. The calculations show that none of these versions is the best in all cases. We can expect that, the combination of these versions will be much more accurate than any of these versions. However, it is not well studied how to determine the prediction for a combination of these single classifiers.

This algorithm is designed for datasets with multi-labeled classes where the prediction is formulated in the form of the order/rank of classes; higher ranked classes are assumed to be more relevant to the tested example under consideration. This type predictions are natural for multi-labeled data classification. For example, SVM and BoosTexter also generate prediction in this form.

Therefore, for the investigation of combinations of such classifiers and their error-correcting capabilities, theorems similar to Theorem 2 will be essential.

## 8 Acknowledgements

The first author acknowledges the support from Australian Research Council, Discovery grant DP0449469. The second author was supported by Queen Elizabeth II Fellowship and ARC Discovery grant DP0211866. The third author has been supported by Discovery grant DP0556685 and Linkage grant LP0669752 from Australian Research Council.

## References

- [1] Alfaro, R. and A.V. Kelarev, *Recent results on ring constructions for error-correcting codes*, “Algebraic Structures and Their Representations”, XV Coloquio Latinoamericano de Algebra (Cocoyoc, Morelos, Mexico, July 20–26, 2003), Contemporary Math. **376** (2005), 1–12.
- [2] Alfaro, R. and A.V. Kelarev, *On cyclic codes in incidence rings*, Studia Sci. Math. Hungarica **43** (2006)(1), 69–77.
- [3] Araújo, I.M., Kelarev, A.V. and A. Solomon, *An algorithm for commutative semigroup algebras which are principal ideal rings with identity*, Comm. Algebra **32** (2004)(4), 1237–1254.
- [4] Ash, C.J., Hall, T.E. and J.-E. Pin, *On the varieties of languages associated with some varieties of finite monoids with commuting idempotents*, Inform. and Comput. **86** (1990)(1), 32–42.
- [5] Auinger, K., Hall, T.E., Reilly, N.R. and S. Zhang, *Congruences on the lattice of pseudovarieties of finite semigroups*, Internat. J. Algebra Comput. **7** (1997)(4), 433–455.
- [6] Cazaran, J. and A.V. Kelarev, *Generators and weights of polynomial codes*, Arch. Math. (Basel) **69** (1997), 479–486.
- [7] Bagirov, A.M. and J.L. Yearwood, *A new nonsmooth optimization algorithm for minimum sum-of-squares clustering problems*, European J. Operational Research **170** (2006), 578–596.
- [8] Bagirov, A.M., Rubinov, A.M. and J. Yearwood, *A global optimization approach to classification*, Optim. Eng. **3** (2002), 129–155.

- [9] Bagirov, A.M., Rubinov, A.M., Soukhoroukova, N.V. and J. Yearwood, *Unsupervised and supervised data classification via nonsmooth and global optimization*, Top **11** (2003), 1–93.
- [10] Cazaran, J., Kelarev, A.V., Quinn, S.J. and D. Vertigan, *An algorithm for computing the minimum distances of extensions of BCH codes embedded in semigroup rings*, Semigroup Forum **73** (2006), 317–329.
- [11] Downey, R., Fellows, M.R., Whittle, G. and A. Vardy, *The parameterized complexity of some fundamental problems in coding theory*, SIAM J. Comput. **29**(2) (1999), 545–570.
- [12] Easdown, D., East, J. and D.G. FitzGerald, *Presentations of factorizable inverse monoids*, Acta Sci. Math. (Szeged) **71** (2005)(3-4), 509–520.
- [13] Easdown, D. and T.G. Lavers, *The inverse braid monoid*, Adv. Math. **186** (2004)(2), 438–455.
- [14] Easdown, D. and L.M. Shneerson, *Principal Rees quotients of free inverse semigroups*, Glasg. Math. J. **45** (2003)(2), 263–267.
- [15] Easdown, D. and W.D. Munn, *Trace functions on inverse semigroup algebras*, Bull. Austral. Math. Soc. **52** (1995)(3), 359–372.
- [16] Ferguson, B., Ghosh, R. and J.L. Yearwood, *Modular neural network design for the problem of alphabetic character recognition*, Internat. J. Pattern Recognition & Artificial Intelligence **19** (2006)(2), 249–269.
- [17] Gomes, G.M.S. and J.M. Howie, *Semigroups with zero whose idempotents form a subsemigroup*, Proc. Roy. Soc. Edinburgh Sect. **A 128** (1998), 265–281.
- [18] Gray, R. and N. Ruškuc, *Generating sets of completely 0-simple semigroups*, Comm. Algebra **33** (2005)(12), 4657–4678.
- [19] Hall, T.E., *The radical of the algebra of any finite semigroup over any field*, J. Austral. Math. Soc. Ser. A **11** (1970), 350–352.
- [20] Hall, T.E., *Biprefix codes, inverse semigroups and syntactic monoids of injective automata*, Theoretical Computer Science **32** (1984), no. 1-2, 201–213.
- [21] Hall, T.E., *Finite inverse semigroups and amalgamation*, “Semigroups and their applications” (Chico, Calif., 1986), 1987, 51–56.
- [22] Hall, T.E., *Amalgamation for inverse and generalized inverse semigroups*, Trans. Amer. Math. Soc. **310** (1988), 313–323.
- [23] Hall, T.E. and T. Imaoka, *Representations and amalgamation of generalized inverse  $*$ -semigroups*, Semigroup Forum **58** (1999), 126–141.
- [24] Hall, T.E. and K.G. Johnston, *The lattice of pseudovarieties of inverse semigroups*, Pacific J. Math. **138** (1989), 73–88.
- [25] Hall, T.E., Kublanovskii, S.I., Margolis, S., Sapir, M.V. and P.G. Trotter, *Algorithmic problems for finite groups and finite 0-simple semigroups*, J. Pure Appl. Algebra **119** (1997)(1), 75–96.



- [26] Hall, T.E. and K. Shoji, *Finite bands and amalgamation bases for finite semigroups*, Comm. Algebra **30** (2002)(2), 911–933.
- [27] Howie, J.M., “Fundamentals of Semigroup Theory”, Clarendon Press, Oxford, 1995.
- [28] Imaoka, T. and T.E. Hall, *Amalgamation of generalized inverse  $\ast$ -semigroups*, Proc. 19th Symposium on Semigroups, Languages and their Related Fields (Matsue, 1995), 8–14, Shimane Univ., Matsue, 1995.
- [29] Joachims, T., “Learning to Classify Text Using Support Vector Machines: Methods, Theory and Algorithms”, The International Series in Engineering and Computer Science, Kluwer / Springer, 2002.
- [30] Kelarev, A.V., “Ring Constructions and Applications”, World Scientific, London, 2002.
- [31] Kelarev, A.V., “Graph Algebras and Automata”, Marcel Dekker, New York, 2003.
- [32] Kelarev, A.V. and D.S. Passman, *A description of incidence rings of group automata*, Contemporary Mathematics **456**(2008), 27–33.
- [33] Lashkarizadeh, B.M. and H. Samea, *Approximate amenability of certain semigroup algebras*, Semigroup Forum **71** (2005)(2), 312–322.
- [34] Luger, G., “Artificial Intelligence. Structures and Strategies for Complex Problem Solving”, 5th edition, Addison-Wesley, 2005.
- [35] Mammadov, M.A., Rubinov, A.M. and J. Yearwood, *The study of drug-reaction relationships using global optimization techniques*, “Optimization Methods and Software”, 2007.
- [36] Margolis, S.W. and J.C. Meakin, *E-unitary inverse monoids and the Cayley graph of a group representation*, J. Pure & Applied Algebra **58** (1989), 45–76.
- [37] Mitchell, J.D., *Turán’s graph theorem and maximum independent sets in Brandt semigroups*, “Semigroups and Languages”, World Sci. Publ., River Edge, NJ, 2004, 151–162.
- [38] Okniński, J., “Semigroup Algebras”, Marcel Dekker, New York, 1991.
- [39] Schapire, R.E. and Y. Singer, *BoosTexter: A boosting-based system for text categorization*, Machine Learning **39**(2000)(2/3), 135–168.
- [40] Shevrin, L.N. and A.Ja. Ovsyannikov, “Semigroups and their Subsemigroup Lattices”, Kluwer, Dordrecht, 1996.
- [41] Tilakaratne, C.D., *Stock Market Prediction Based on Quantified Intermarket Influence*, PhD Thesis, University of Ballarat, Australia, 2007.
- [42] Yearwood, J.L. and M. Mammadov, “Classification Technologies: Optimization Approaches to Short Text Categorization”, Idea Group Inc., 2007.
- [43] Yearwood, J.L. and A. Stranieri, *The generic/actual argument model of practical reasoning*, Decision Support Systems **41** (2006), 358–379.
- [44] Witten, I.H. and E. Frank, “Data Mining: Practical Machine Learning Tools and Techniques”, Elsevier/Morgan Kaufman, Amsterdam, 2005.